



FY 5620  
#12

PCT

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/807,933B

DATE: 01/22/2003

TIME: 13:30:49

Input Set : A:\09807933SeqList.txt

Output Set: N:\CRF4\01222003\I807933B.raw

4 <110> APPLICANT: MEIJI SEIKA KAISHA, LTD.  
 6 <120> TITLE OF INVENTION: Endoglucanase and cellulase composition containing the  
 7 same  
 9 <130> FILE REFERENCE: 121659PX  
 C--> 11 <140> CURRENT APPLICATION NUMBER: US/09/807,933B  
 C--> 12 <141> CURRENT FILING DATE: 2002-04-18  
 14 <150> PRIOR APPLICATION NUMBER: JP302387/1998  
 15 <151> PRIOR FILING DATE: 1998-10-23  
 17 <160> NUMBER OF SEQ ID NOS: 113  
 19 <170> SOFTWARE: PatentIn Ver. 2.0  
 21 <210> SEQ ID NO: 1  
 22 <211> LENGTH: 338  
 23 <212> TYPE: PRT  
 24 <213> ORGANISM: Rhizopus oryzae CP96001  
 26 <220> FEATURE:  
 27 <221> NAME/KEY: sig\_peptide  
 28 <222> LOCATION: (-23)..(-1)  
 29 <221> NAME/KEY: mat\_peptide  
 30 <222> LOCATION: (1)..(315)  
 W--> 32 <400> 1  
 33 Met Lys Phe Ile Thr Ile Ala Ser Ser Ala Leu Leu Ala Leu Ala Leu  
 34 -20 -15 -10  
 36 Gly Thr Glu Met Ala Ser Ala Ala Glu Cys Ser Lys Leu Tyr Gly Gln  
 37 -5 1 5  
 39 Cys Gly Gly Lys Asn Trp Asn Gly Pro Thr Cys Cys Glu Ser Gly Ser  
 40 10 15 20 25  
 42 Thr Cys Lys Val Ser Asn Asp Tyr Tyr Ser Gln Cys Leu Pro Ser Gly  
 43 30 35 40  
 45 Ser Ser Gly Asn Lys Ser Ser Glu Ser Ala His Lys Lys Thr Thr Thr  
 46 45 50 55  
 48 Ala Ala His Lys Lys Thr Thr Thr Thr Ala Ala His Lys Lys Thr Thr Thr  
 49 60 65 70  
 51 Ala Pro Ala Lys Lys Thr Thr Thr Val Ala Lys Ala Ser Thr Pro Ser  
 52 75 80 85  
 54 Asn Ser Ser Ser Ser Ser Ser Gly Lys Tyr Ser Ala Val Ser Gly Gly  
 55 90 95 100 105  
 57 Ala Ser Gly Asn Gly Val Thr Thr Arg Tyr Trp Asp Cys Cys Lys Ala  
 58 110 115 120  
 60 Ser Cys Ser Trp Pro Gly Lys Ala Asn Val Ser Ser Pro Val Lys Ser  
 61 125 130 135  
 63 Cys Asn Lys Asp Gly Val Thr Ala Leu Ser Asp Ser Asn Ala Gln Ser  
 64 140 145 150  
 66 Gly Cys Asn Gly Gly Asn Ser Tyr Met Cys Asn Asp Asn Gln Pro Trp

ENTERED

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/807,933B

DATE: 01/22/2003

TIME: 13:30:49

Input Set : A:\09807933SeqList.txt

Output Set: N:\CRF4\01222003\I807933B.raw

```

67      155      160      165
69 Ala Val Asn Asp Asn Leu Ala Tyr Gly Phe Ala Ala Ala Ala Ile Ser
70 170      175      180      185
72 Gly Gly Gly Glu Ser Arg Trp Cys Cys Ser Cys Phe Glu Leu Thr Phe
73      190      195      200
75 Thr Ser Thr Ser Val Ala Gly Lys Lys Met Val Val Gln Val Thr Asn
76      205      210      215
78 Thr Gly Gly Asp Leu Gly Ser Ser Thr Gly Ala His Phe Asp Leu Gln
79      220      225      230
81 Met Pro Gly Gly Gly Val Gly Ile Phe Asn Gly Cys Ser Ser Gln Trp
82      235      240      245
84 Gly Ala Pro Asn Asp Gly Trp Gly Ser Arg Tyr Gly Gly Ile Ser Ser
85 250      255      260      265
87 Ala Ser Asp Cys Ser Ser Leu Pro Ser Ala Leu Gln Ala Gly Cys Lys
88      270      275      280
90 Trp Arg Phe Asn Trp Phe Lys Asn Ala Asp Asn Pro Ser Met Thr Tyr
91      285      290      295
93 Lys Glu Val Thr Cys Pro Lys Glu Ile Thr Ala Lys Thr Gly Cys Ser
94      300      305      310
96 Arg Lys
97      315

```

99 &lt;210&gt; SEQ ID NO: 2

100 &lt;211&gt; LENGTH: 1017

101 &lt;212&gt; TYPE: DNA

102 &lt;213&gt; ORGANISM: Rhizopus oryzae CP96001

104 &lt;220&gt; FEATURE:

105 &lt;221&gt; NAME/KEY: sig\_peptide

106 &lt;222&gt; LOCATION: (1)..(69)

107 &lt;221&gt; NAME/KEY: mat\_peptide

108 &lt;222&gt; LOCATION: (70)..(1017)

W--&gt; 110 &lt;400&gt; 2

```

111 atg aag ttt att act att gcc tct tcc gct ctc ttg gct ctc gcc ctc 48
112 Met Lys Phe Ile Thr Ile Ala Ser Ser Ala Leu Leu Ala Leu Ala Leu
113      -20      -15      -10
115 ggt act gaa atg gcc tct gct gct gaa tgt agc aaa ttg tat ggt caa 96
116 Gly Thr Glu Met Ala Ser Ala Ala Glu Cys Ser Lys Leu Tyr Gly Gln
117      -5      1      5
119 tgt ggt ggt aag aac tgg aat ggc cct act tgt tgt gaa tct gga tcc 144
120 Cys Gly Gly Lys Asn Trp Asn Gly Pro Thr Cys Cys Glu Ser Gly Ser
121 10      15      20      25
123 acc tgt aaa gta agc aac gat tac tac tct caa tgt ctt ccc tct gga 192
124 Thr Cys Lys Val Ser Asn Asp Tyr Tyr Ser Gln Cys Leu Pro Ser Gly
125      30      35      40
127 agc agt ggc aat aaa tct tct gaa agt gct cac aag aag act acc act 240
128 Ser Ser Gly Asn Lys Ser Ser Glu Ser Ala His Lys Lys Thr Thr Thr
129      45      50      55
131 gct gct cac aag aag act act acc gct gct cat aaa aag act acc act 288
132 Ala Ala His Lys Lys Thr Thr Thr Ala Ala His Lys Lys Thr Thr Thr
133      60      65      70

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/807,933B

DATE: 01/22/2003

TIME: 13:30:49

Input Set : A:\09807933SeqList.txt

Output Set: N:\CRF4\01222003\I807933B.raw

```

135 gct cct gct aag aag act aca act gtt gcc aaa gct tcc acc cct tct 336
136 Ala Pro Ala Lys Lys Thr Thr Thr Val Ala Lys Ala Ser Thr Pro Ser
137 75 80 85
139 aac tct agc tct agc tcc agc ggc aaa tat tcc gct gtc tct ggt ggt 384
140 Asn Ser Ser Ser Ser Ser Ser Gly Lys Tyr Ser Ala Val Ser Gly Gly
141 90 95 100 105
143 gcc tct ggt aac ggt gtc act act cgt tat tgg gat tgc tgt aag gcc 432
144 Ala Ser Gly Asn Gly Val Thr Thr Arg Tyr Trp Asp Cys Cys Lys Ala
145 110 115 120
147 tcc tgt agc tgg ccc ggt aag gcc aat gtc agt tct cct gtc aag tcc 480
148 Ser Cys Ser Trp Pro Gly Lys Ala Asn Val Ser Ser Pro Val Lys Ser
149 125 130 135
151 tgt aac aaa gat ggt gtc act gcc ctt agt gac agc aat gcc caa agt 528
152 Cys Asn Lys Asp Gly Val Thr Ala Leu Ser Asp Ser Asn Ala Gln Ser
153 140 145 150
155 ggc tgt aac ggt ggt aac agt tac atg tgt aac gac aac caa cct tgg 576
156 Gly Cys Asn Gly Gly Asn Ser Tyr Met Cys Asn Asp Asn Gln Pro Trp
157 155 160 165
159 gct gta aac gac aac ctt gcc tat ggt ttc gct gct gct gcc atc agt 624
160 Ala Val Asn Asp Asn Leu Ala Tyr Gly Phe Ala Ala Ala Ala Ile Ser
161 170 175 180 185
163 ggt ggt ggt gaa tct cgc tgg tgc tgt tct tgt ttc gaa ctt act ttc 672
164 Gly Gly Gly Glu Ser Arg Trp Cys Cys Ser Cys Phe Glu Leu Thr Phe
165 190 195 200
167 act tct acc tct gtt gct ggt aag aag atg gtt gtc caa gtc act aac 720
168 Thr Ser Thr Ser Val Ala Gly Lys Lys Met Val Val Gln Val Thr Asn
169 205 210 215
171 act ggt ggt gat ctt ggc tcc tct act ggt gct cac ttt gac ttg caa 768
172 Thr Gly Gly Asp Leu Gly Ser Ser Thr Gly Ala His Phe Asp Leu Gln
173 220 225 230
175 atg ccc ggt ggt ggt gtt ggt att ttc aat ggt tgt tcc agc caa tgg 816
176 Met Pro Gly Gly Gly Val Gly Ile Phe Asn Gly Cys Ser Ser Gln Trp
177 235 240 245
179 ggt gct ccc aat gac ggt tgg ggc tca aga tac ggt ggt att tct tct 864
180 Gly Ala Pro Asn Asp Gly Trp Gly Ser Arg Tyr Gly Gly Ile Ser Ser
181 250 255 260 265
183 gca tct gac tgc tct agt ctt cct tcc gca ctc caa gct ggt tgt aaa 912
184 Ala Ser Asp Cys Ser Ser Leu Pro Ser Ala Leu Gln Ala Gly Cys Lys
185 270 275 280
187 tgg aga ttc aac tgg ttc aag aac gct gat aac cca agc atg act tac 960
188 Trp Arg Phe Asn Trp Phe Lys Asn Ala Asp Asn Pro Ser Met Thr Tyr
189 285 290 295
191 aag gaa gtt acc tgt cct aag gaa atc acc gcc aag aca ggt tgt tca 1008
192 Lys Glu Val Thr Cys Pro Lys Glu Ile Thr Ala Lys Thr Gly Cys Ser
193 300 305 310
195 aga aaa taa 1017
196 Arg Lys
197 315
199 <210> SEQ ID NO: 3

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/807,933B

DATE: 01/22/2003

TIME: 13:30:49

Input Set : A:\09807933SeqList.txt

Output Set: N:\CRF4\01222003\I807933B.raw

200 &lt;211&gt; LENGTH: 366

201 &lt;212&gt; TYPE: PRT

202 &lt;213&gt; ORGANISM: Rhizopus oryzae CP96001

204 &lt;220&gt; FEATURE:

205 &lt;221&gt; NAME/KEY: sig\_peptide

206 &lt;222&gt; LOCATION: (-23)..(-1)

207 &lt;221&gt; NAME/KEY: mat\_peptide

208 &lt;222&gt; LOCATION: (1)..(343)

W--&gt; 210 &lt;400&gt; 3

```

211 Met Lys Phe Ile Thr Ile Thr Ser Ser Ala Leu Leu Ala Leu Ala Leu
212          -20                      -15                      -10
214 Gly Thr Glu Met Ala Ser Ala Ala Lys Cys Ser Lys Leu Tyr Gly Gln
215          -5                      1                      5
217 Cys Gly Gly Lys Asp Trp Asn Gly Pro Thr Cys Cys Glu Ser Gly Ser
218 10          15                      20                      25
220 Thr Cys Lys Val Ser Asn Asp Tyr Tyr Ser Gln Cys Leu Ala Pro Glu
221          30                      35                      40
223 Ser Asn Gly Asn Lys Ser Ser Glu Cys Ser Lys Leu Tyr Gly Gln Cys
224          45                      50                      55
226 Gly Gly Lys Asp Trp Asn Gly Pro Thr Cys Cys Glu Ser Gly Ser Thr
227          60                      65                      70
229 Cys Lys Val Ser Asn Asp Tyr Tyr Ser Gln Cys Leu Ala Pro Glu Ser
230          75                      80                      85
232 Asn Gly Asn Lys Thr Ser Glu Ser Ala His Lys Thr Thr Thr Thr Thr
233 90          95                      100                      105
235 Ala Pro Ala Lys Glu Ile Thr Thr Thr Ala Lys Ala Ser Asn Ser Ser
236          110                      115                      120
238 Asn Ser Ser Gly Lys Tyr Ser Ile Val Ser Gly Gly Ala Ser Gly Asn
239          125                      130                      135
241 Gly Val Thr Thr Arg Tyr Trp Asp Cys Cys Lys Ala Ser Cys Ser Trp
242          140                      145                      150
244 Pro Gly Lys Lys Ala Asn Val Ser Ser Pro Val Lys Ser Cys Asn Lys Asp
245          155                      160                      165
247 Gly Val Thr Ala Leu Ser Asp Ser Asn Val Gln Ser Gly Cys Asn Gly
248 170          175                      180                      185
250 Gly Asn Ser Tyr Met Cys Asn Asp Asn Gln Pro Trp Ala Val Asn Asp
251          190                      195                      200
253 Asn Leu Ala Tyr Gly Phe Ala Ala Ala Ile Ser Gly Gly Gly Glu
254          205                      210                      215
256 Ser Arg Trp Cys Cys Ser Cys Phe Glu Leu Thr Phe Thr Ser Thr Ser
257          220                      225                      230
259 Val Ala Gly Lys Lys Met Val Ile Gln Val Thr Asn Thr Gly Gly Asp
260          235                      240                      245
262 Leu Gly Ser Ser Thr Gly Ala His Phe Asp Leu Gln Met Pro Gly Gly
263 250          255                      260                      265
265 Gly Val Gly Ile Phe Asn Gly Cys Ser Lys Gln Trp Gly Ala Pro Asn
266          270                      275                      280
268 Asp Gly Trp Gly Ser Arg Tyr Gly Gly Ile Ser Ser Ala Ser Asp Cys
269          285                      290                      295

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/807,933B

DATE: 01/22/2003

TIME: 13:30:49

Input Set : A:\09807933SeqList.txt

Output Set: N:\CRF4\01222003\I807933B.raw

271 Ser Ser Leu Pro Ser Ala Leu Gln Ala Gly Cys Lys Trp Arg Phe Asn  
 272 300 305 310  
 274 Trp Phe Lys Asn Ala Asp Asn Pro Ser Met Thr Tyr Lys Glu Val Thr  
 275 315 320 325  
 277 Cys Pro Lys Glu Ile Thr Ala Lys Thr Gly Cys Ser Arg Lys  
 278 330 335 340

280 &lt;210&gt; SEQ ID NO: 4

281 &lt;211&gt; LENGTH: 1101

282 &lt;212&gt; TYPE: DNA

283 &lt;213&gt; ORGANISM: Rhizopus oryzae CP96001

285 &lt;220&gt; FEATURE:

286 &lt;221&gt; NAME/KEY: sig\_peptide

287 &lt;222&gt; LOCATION: (1)..(69)

288 &lt;221&gt; NAME/KEY: mat\_peptide

289 &lt;222&gt; LOCATION: (70)..(1101)

W--&gt; 291 &lt;400&gt; 4

292 atg aag ttt att act att acc tct tcc gct ctc ttg gct ctc gcc ctt 48  
 293 Met Lys Phe Ile Thr Ile Thr Ser Ser Ala Leu Leu Ala Leu Ala Leu  
 294 -20 -15 -10  
 296 ggt act gaa atg gcc tct gct gct aaa tgt agc aag ctg tat ggt caa 96  
 297 Gly Thr Glu Met Ala Ser Ala Ala Lys Cys Ser Lys Leu Tyr Gly Gln  
 298 -5 1 5  
 300 tgt ggt ggt aag gac tgg aat ggc cct act tgt tgc gaa tct gga tcc 144  
 301 Cys Gly Gly Lys Asp Trp Asn Gly Pro Thr Cys Cys Glu Ser Gly Ser  
 302 10 15 20 25  
 304 acc tgt aaa gta agc aac gat tac tac tct caa tgt ctt gcc cct gaa 192  
 305 Thr Cys Lys Val Ser Asn Asp Tyr Tyr Ser Gln Cys Leu Ala Pro Glu  
 306 30 35 40  
 308 agc aac ggc aat aag tct tct gaa tgt agc aag ttg tat ggt caa tgt 240  
 309 Ser Asn Gly Asn Lys Ser Ser Glu Cys Ser Lys Leu Tyr Gly Gln Cys  
 310 45 50 55  
 312 ggt ggt aag gac tgg aat ggc cct act tgt tgc gaa tct gga tcc acc 288  
 313 Gly Gly Lys Asp Trp Asn Gly Pro Thr Cys Cys Glu Ser Gly Ser Thr  
 314 60 65 70  
 316 tgt aaa gta agc aac gat tac tac tct caa tgt ctt gcc cct gaa agc 336  
 317 Cys Lys Val Ser Asn Asp Tyr Tyr Ser Gln Cys Leu Ala Pro Glu Ser  
 318 75 80 85  
 320 aat ggc aat aaa act tct gaa agc gct cat aaa acg act act acc act 384  
 321 Asn Gly Asn Lys Thr Ser Glu Ser Ala His Lys Thr Thr Thr Thr Thr  
 322 90 95 100 105  
 324 gct ccc gct aag gaa att aca act act gcc aaa gct tca aac tct tct 432  
 325 Ala Pro Ala Lys Glu Ile Thr Thr Thr Ala Lys Ala Ser Asn Ser Ser  
 326 110 115 120  
 328 aac tct agc ggc aaa tac tcc att gtc tct ggt ggt gcc tct ggt aac 480  
 329 Asn Ser Ser Gly Lys Tyr Ser Ile Val Ser Gly Gly Ala Ser Gly Asn  
 330 125 130 135  
 332 ggt gtc act act cgt tat tgg gat tgc tgt aag gcc tcc tgt agc tgg 528  
 333 Gly Val Thr Thr Arg Tyr Trp Asp Cys Cys Lys Ala Ser Cys Ser Trp  
 334 140 145 150

RAW SEQUENCE LISTING ERROR SUMMARY  
PATENT APPLICATION: US/09/807,933B

DATE: 01/22/2003  
TIME: 13:30:50

Input Set : A:\09807933SeqList.txt  
Output Set: N:\CRF4\01222003\I807933B.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:17; Xaa Pos. 1,2,3,4,5,6,11,12,13,14,16,17,18,20,21,22,23,24,25,27,28  
Seq#:17; Xaa Pos. 29,30,31,33,34,36,39  
Seq#:18; Xaa Pos. 1,2,3,4,5,6,11,12,13,14,16,17,18,20,21,22,23,24,25,27,28  
Seq#:18; Xaa Pos. 29,30,31,33,34,36,39  
Seq#:19; Xaa Pos. 3,4,6,11,12,14,20,21,23,26,27,28,29,30,32,33,38  
Seq#:20; Xaa Pos. 3,4,6,11,12,14,20,21,23,26,27,28,29,30,32,33,38  
Seq#:21; Xaa Pos. 4,6,12,14,26,27,28,29,30,32,33,38  
Seq#:25; Xaa Pos. 6,14,20,21,26,29,30,38  
Seq#:29; Xaa Pos. 1,5,7,8,9,10,11,12  
Seq#:30; Xaa Pos. 2,3,4,8  
Seq#:31; Xaa Pos. 2,3,4,5,8,9  
Seq#:32; Xaa Pos. 2,3,4,5,8,9  
Seq#:43; N Pos. 15,18  
Seq#:44; N Pos. 15